Docket No. MCD118 CIP
Client/Matter No. 83629.0037.001

METHOD FOR SELECTIVELY TRANSMITTING FRAMES THROUGH A SWITCH ACCORDING TO A QUALITY OF SERVICE

RELATED APPLICATIONS

This application is a continuation-in-part of co-pending U.S. Patent Application Serial No. 09/728,452 filed November 30, 2000, which is a continuation-in-part of U.S. Patent Application Serial No. 09/228,678 filed January 12, 1999 (now U.S. Patent No. 6,233,236), both of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

Field Of The Invention

This invention pertains generally to improvements in methods for sequencing data through a routing device. More particularly, this invention pertains to a method for scoring queued frames for selective transmission through a switch. This invention is particularly, but not exclusively, useful for selectively transmitting frames through a fibre channel switch.

Relevant Background

Computer performance during the past decade has increased significantly, if not exponentially, in part driven by the information explosion. Consequently, demand for high-performance communications for server-to-storage and server-to-server networking also has increased. Performance improvements in hardware entities, including storage, processors, and workstations, along with the move to distributed architectures such as client/server, have increased the demand for data-intensive and high-speed networking applications. The interconnections between and among these systems, and their input/output devices, require enhanced levels of performance in reliability, speed, and distance.

Simultaneously, demands for more robust, highly available, disaster-tolerant computing resources, with ever-increasing speed and memory capabilities, continues unabated. To satisfy such demands, the computer industry has worked to overcome performance problems often attributable to conventional I/0 ("input/output devices") subsystems. Mainframes, supercomputers, mass storage systems, workstations and very high-resolution display subsystems frequently are connected to facilitate file and print sharing. Because of the demand for increased speed across such systems, networks and channels conventionally used for connections introduce communication clogging, especially if data is in large file format typical of graphically based applications.

Efforts to satisfy an enhanced performance demands have been directed to providing storage interconnect solutions that address performance and reliability requirements of modern storage systems. At least three technologies are directed to solving those problems SCSI ("Small Computer Systems Interface"); SSA ("Serial Storage Architecture"), a technology advanced primarily by IBM; and Fibre Channel, a high performance interconnect technology.

Two prevalent types; of data communication connections exist between processors, and between a processor and peripherals. A "channel" provides direct or switched point-to-point connection communicating devices. The channels primary task is to transport data at the highest possible data rate with the least amount o delay. Channels typically perform simple error correction in hardware. A "network", by contrast, is an aggregation of distributed nodes. A "node" as used in this document is either an individual computer or similar machine in a network (workstations, mass storage units, etc.) with a protocol that supports interaction among the nodes. Typically, each node must be capable of recognizing error conditions on the network and must provide the error management required to recover from the error conditions.

SCSI is an "intelligent" and parallel I/O bus on which various peripheral devices and controllers can exchange information. Although designed approximately 15 years ago, SCSI remains in use. The first SCSI standard, now known as SCSI-1, was adopted in 1986 and originally designed to accommodate up to eight devices at speeds of 5 MB/sec. SCSI standards and technology have been refined and extended frequently, providing ever faster data: transfer rates up to 40MB/sec. SCSI performance has doubled approximately every five years since the original standard was released, and the number of devices permitted on a single bus has been increased to 16. In addition, backward compatibility has been enhanced, enabling newer devices to coexist on a bus with older devices. However, significant problems associated with SCSI remain, including, for example, limitations caused by bus speed, bus length, reliability, cost, and device count. In connection with bus length, originally limited to six meters, newer standards requiring even faster transfer rates and higher device populations now place more stringent limitations on bus length that are only partially cured by expensive differential cabling or extenders.

Accordingly, industry designers now seek to solve the limitations inherent in SCSI by employing serial device interfaces. Featuring data transfer rates as high as 200 MB/sec, serial interfaces use point-to-point interconnections rather than busses. Serial designs also decrease cable complexity, simplify electrical requirements, and increase reliability. Two solutions have been considered, Serial Storage Architecture and what has become known as Fibre Channel technology, including the Fibre Channel Arbitrated Loop`("FC-AL").

Serial Storage Architecture is a high-speed serial interface designed to connect data storage devices, subsystem, servers and workstations. SSA was developed and is promoted as an industry standard by IBM; formal standardization processes began in 1992. Currently, SSA is undergoing approval processes as an ANSI standard. Although the basic transfer rate through an SSA port is only 20 MB/sec, SSA is dual ported and full-duplex, resulting in a maximum aggregate transfer speed of up to 80 MB/sec. SSA connections are carried over thin, shielded, four-wire (two differential pairs) cables, which are less expensive and more flexible than the typical 50- and 68-conductor SCSI cables. Currently, IBM is the only major disk drive manufacturer shipping SSA drives; there has been little industry-wide support for SSA. That is not true of Fibre Channel, which has achieved wide industry support.

Fibre Channel ("F/C") is an industry-standard, high-speed serial data transfer interface used to connect systems and storage in point-to-point or switched topologies. FC-AL technology, developed with storage connectivity in mind, is a recent enhancement that also supports copper media and loops containing up to 126 devices, or nodes. Briefly, fibre channel is a switched protocol that allows concurrent communication among workstations, super computers and various peripherals. The total network bandwidth provided by fibre channel may be on the order of a terabit per second. Fibre channel is capable of transmitting frames along lines or lanes at rates exceeding 1 gigabit per second in at least two directions simultaneously. F/C technology also is able to transport commands and data according to existing protocols such a Internet protocol ("IP"), high performance parallel interface ("HIPPI"), intelligent peripheral interface ("IPI"), and, as indicated using SCSI, over and across both optical fiber and copper cable.

The fibre channel maybe considered a channel-network hybrid. An F/C system contains sufficient network features to provide connectivity, distance and protocol multiplexing, and enough channel features to retain simplicity, repeatable performance and reliable delivery. Fibre channel allows for an active, intelligent interconnection scheme, known as a "fabric", and fibre channel switches to connect devices. The F/C fabric includes a plurality of fabric-ports (F_ports) that provide for interconnection and frame transfer between plurality of node-ports (N_ports) attached to associated devices that may include workstations, super computers and/or peripherals. A fabric has the capability of routing frames based on information contained within the frames. The N_port transmits and receives data to and from the fabric.

Transmission is isolated from the control protocol so that different topologies (e.g., point-to-point links, rings, multidrop buses, and crosspoint switches) can be implemented. Fibre Channel, a highly reliable, gigabit interconnect technology allows concurrent communications among workstations, mainframes, servers, data storage systems, and other peripherals. F/C technology provides interconnect systems for multiple topologies that can scale to a total system bandwidth on the order of a terabit per second. Fibre Channel delivers a new level of reliability and throughput. Switches, hubs, storage systems, storage devices, and adapters designed for the F/C environment are available now.

Following a lengthy review of existing equipment and standards, the Fibre Charnel standards group realized that it would be useful for channels and networks to share the same fiber. (The terms "fiber' or "fibre" are used synonymously, and include both optical and copper cables.) The Fibre Channel protocol was developed and adopted, and continues to be developed, as the American National Standard for Information Systems ("ANSI"). See Fibre Channel Physical and Signaling Interface, Revision 4.2, American National Standard for Information Systems (ANSI) (1993) for a detailed discussion of the fibre channel standard, which is incorporated by reference into this document.

Current standards for F/C support bandwidth of 133 Mb/sec, 266 Mb/sec, 532 Mb/sec, 1.0625 Gb/sec, and 4 Gb/sec (proposed) at distances of up to ten kilometers. Fibre Channel's current maximum data rate is 100 MB/sec (200 MB/sec full-duplex) after accounting for overhead. In addition to strong channel characteristics, Fibre Channel provides powerful networking capabilities, allowing switches and hubs to interconnect systems and storage into tightly knit clusters. The clusters are capable of providing high levels of performance for file service, database management, or general purpose computing. Because Fibre Channel is able to span up to 10 kilometers between nodes, F/C allows very high-speed movement of data between systems that are greatly separated from one another. The F/C standard defines a layered protocol architecture consisting of five layers, the highest layer defining mappings from other communication protocols onto the F/C fabric.

The network behind the servers link one or more servers to one or more storage systems. Each storage system could be RAID ("Redundant Array of Inexpensive Disks"), tape backup, tape library, CD-ROM library, or JBOD

("Just a Bunch of Disks"). One type of RAID system divides each byte of data into bits and stores each bit on a different disk. If the data consists of 8-bit bytes, there will be 10 disks, one for each of the 8 bits, and two more for an error-correcting code. The error-correcting code makes it possible to reconstruct any single missing bit in any byte. Thus, if one of the disk drives fails completely, only one bit will be missing from each byte, and the contents of the failed disk can be reconstructed completely from the error-correcting code.

Fibre Channel networks have proven robust and resilient, and include at least these features: shared storage among systems; scalable networking; high performance; fast data access and backup. In a Fibre Channel network, legacy storage systems are interfaced using a Fibre Channel to SCSI bridge. Fibre Channel standards include network features that provide required connectivity, distance, and protocol multiplexing. It also supports traditional channel features for simplicity, repeatable performance, and guaranteed delivery.

The demand for speed and volume of transmission has generated a concomitant demand for a capability to sort data to enable a user to identify data and data streams that have higher priority than other data queued in devices for routing data, such as a switch, particularly a fibre channel switch. It would be useful, therefore, to be able to order, or sequence, transmission of data through a fibre channel switch, including frames, based on the content of the frame as well as the source of the frame by assigning a score to data received by a device such as a switch, and to be able to transmit data and frames having the highest score.

Currently, therefore, a previously unaddressed need exists in the industry for new, useful and reliable method for scoring queued frames for selective transmission through a switch, particularly in a Fibre Channel environment. It would be of considerable advantage to provide a method for assigning scores to data frames received by a switch, and to selectively expedite transmission of the frames having the highest score.

SUMMARY OF THE INVENTION

In accordance with the present invention a method for scoring queued frames for selective transmission through a routing device, including a switch, is provided. The invention provides for one or more fibre channel switches. The invention also includes the receipt of data, including frames, by the one or more fibre channel switches at a connection, particularly a receiving port in the fibre channel switch. The switches, particularly fibre channel switches, are equipped with one or more registers. Further, the fibre channel switches include one or more means for programming the registers. The data may be received in any order, in sequence or not in sequence.

On receipt of the data, including one or more frames, the data is evaluated based on the content of the data, and an initial score is assigned to the content of the data, including one or more frames of data. The initial score is assigned to the data using a quality of service value based on the content of the data. At least one way to determine the initial value using a quality of service value based on content of the data is to locate at a specific location, on each of the frames, data that has been chosen for examination or scoring, information that is conveyed to the registers from the means for programming the registers associated with the switch. Means for programming the registers are included in all fibre channel switches, and may include a range of software programmability options. Also included in the means for programming are data templates that may be used to program the registers for specific purposes, and may be used for examining predetermined data, or data selected for high priority transmission through the switch. The data thus selected may be bit-wise ANDed with a data mask to obtain revised data. The revised data may be compared with the predetermined data to determine if a participation match exists. Alternatively, the adjusting step also permits selecting frames for which there is no participation match.

The present invention also includes steps for adjusting the initial score. A variety of alternative score components may be applied to the initial score to determine one or more adjusted scores. For example, at least one embodiment of the present invention includes a step for identifying the connections, or receiving ports, in the switch where frames are received. A bandwidth allocation adjustment may be applied to the initial score, an adjustment derived from the data location among the receiving ports. In addition, the present invention provides for calculating the cumulative time that the data, including frames, remain on queue in transmit switches of the switch before being transmitted to a receiver device in the fibre channel fabric. The invention provides for measuring each millisecond (ms) of data time on queue, but different time intervals may be selected. Generally, the initial score is increased by the bandwidth allocation adjustment and for the time on queue.

The present invention also includes a step by which the adjusted scores may be compared with each other. Frames having the highest adjusted scores are identified, and the data having the highest score is rearranged in a reordered queue based on the adjusted scores. The frames having the highest adjusted scores may then be transmitted through the switches.

The foregoing has outlined broadly the more important features of the invention to better understand the detailed description that follows, and to better understand the contribution of the present invention to the art. Before explaining at least one embodiment of the invention in detail, it is to be understood that the invention is not limited in application to the details of construction, and to the arrangements of the components, provided in the following description or drawing figures. The invention is capable of other embodiments, and of being practiced and carried out in various ways. In addition, the phraseology and terminology, employed in this disclosure are for purpose of description, and should not be regarded as limiting.

At least one advantage of the present invention is that it enhances the availability of the delivery of data frames having higher priority than other frames.

Another advantage o the present invention is that it provides a method for scoring queued frames in a switch for selective transmission through the switch using programmable elements of fibre channel switches already known in the industry.

The present invention also will permit flexibility in selecting among alternative score components to assign scores to the frames received by a fibre channel switch.

Yet another advantage of the present invention is a method for selectively transmitting frames across a fibre channel fabric that is easy to use and to practice, and is cost effective.

These advantages, and other objects and features, of such a method for scoring queued frames for selective transmission through a routing device, including a switch, will become apparent to those skilled in the art when read in conjunction with the accompanying following description, drawing figures, and appended claims.

As those skilled in the art will appreciate, the conception on which this disclosure is based readily may be used as a basis for designing other structures, methods, and systems for carrying out the purposes of the present invention. The claims, therefore, include such equivalent constructions to the extent the equivalent constructions do not depart from the spirit and scope of the present invention. Further, the abstract associated with this disclosure is neither intended to define the invention, which is measured by the claims, nor intended to be limiting as to the scope of the invention in any way.

BRIEF DESCRIPTION OF THE DRAWING

The novel features of this invention, and the invention itself, both as to structure and operation, are best understood from the accompanying drawing, considered in connection with the accompanying description of the drawing, in which similar reference characters refer to similar parts, and in which:

Figure 1 is a schematic block diagram showing the steps in the method for scoring queued frames for selective transmission through a switch;

Figure 2 is a perspective view showing one of many ways a number of devices, including a Fibre Channel switch, may be interconnected in a Fibre Channel network;

Figure 3 is schematic representation of a variable-length frame communicated through a fiber optic switch as contemplated by the Fibre Channel industry standard;

Figure 4 is a schematic block flow diagram showing one way a quality of service value in accordance with the present invention may be accomplished; and

Figure 5 is schematic block diagram showing one way the method of the present invention may assign a score.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Briefly, the present invention provides a method for scoring queued frames for selective transmission through a switch, particularly a fibre channel switch. As used in this document, the term "scoring" describes the objective of the present invention, namely to assign a score to selected data, including data on a frame, and to select scores having the highest value for priority transmission across devices, including a switch, in a fibre channel fabric. The terms "queued" or "queue" refers to one or more data structures from which items are removed, but for application of the present invention, in the same order in which they were entered. A "frame," as used in this document, includes a presumed configuration of an aggregation of data bits as exemplified in Figure 3.

As shown in Figure 1, the invention generally includes providing one or more switches 12, particularly a fibre channel switch 12' in a fibre channel fabric generally shown in Figure 2; receiving data, including frames 18, as exemplified in Figure 3, at connections that include ports 16 in switches 12 as described in this document; assigning an initial score 20 to the content of the one or more frames 18; adjusting the initial score 22, as shown in Figure 1, with one or more alternative score components to determine one or more adjusted scores; comparing the adjusted scores; selecting frames having the highest adjusted scores 24; and transmitting through fibre channel switches 12' frames 18 having the highest adjusted scores.

The present invention, therefore, is useful for enhancing delivery of data frames having higher priority than other frames. The present invention also is useful for scoring queued frames in a switch, particularly a fibre channel switch, for selective transmission through the switch using programmable elements of fibre channel switches already known in the industry. Flexibility in selecting among alternative score components also is included in the present invention.

Referring first to Figure 1, a schematic and block diagram is shown illustrating in general the method for scoring queued frames for selective transmission through a switch, and is generally designated 10. As shown, the method for scoring queued frames for selective transmission through a switch 10 includes providing one or more fibre channel switches 12 as shown best in Figure 2. At least one example of a switch 12 in which the present invention may successfully operate is a fibre channel switch employing distributed source and destination queuing for interconnecting a plurality of devices 14a-f, which may also include workstations, supercomputers, and other peripherals, through their associated node ports. Fibre channel switch 12' provides a fabric having a shared memory coupled to a plurality of fabric ports 16 through a bit-slicing memory controller (not shown) over which memory addresses, frame data and communications commands are transmitted. More particularly, at least one example of a fibre channel switch in which the present invention may successfully operate is described and shown in U.S. Patent No. 5,894,481 issued on April 13, 1999 to Book, a patent that is incorporated by reference into this document.

As shown by cross-reference between Figures 1 and 3, the present invention also includes the receipt of data, including one or more frames 18, by one or more fibre channel switches 12' at a connection, particularly a receiving port 16' in fibre channel switch 12'. Fibre channel switch 12' is equipped with one or more registers (not shown). Although not shown in the accompanying drawing figures, the term "register" or "registers" as used in this document includes at least one or more row of electronic circuits that can switch back and forth between two states (0 and 1), will remain in either state until changed, and are used to store one or more groups of binary digits while a CPU is processing them. Further, as known by those skilled in the art, fibre channel switches 12' include one or more means for programming the registers.

As shown in Figure 3, frames 18 may be received in any order, in sequence or not in sequence. On receipt of the one or more frames, the data on the frames is evaluated based on the content of the data, and an initial score is assigned to the content of the one or more frames 18. More particularly, as shown in Figure 4, the initial score is assigned to data on frame 18 using a quality of service value 24 based on the content of the data. At least one way to determine the initial value using a quality of service value based on content is to locate, at a specific location on each frame 18, data 26, as exemplified in Figure 3, that has been chosen for examination. Quality of service value 24 is communicated to the registers from the means for programming (not shown) included in the fibre channel switch 12'. The means for programming may include software. Included in the means for programming are data templates that maybe used to program the registers and used for examining the predetermined data 26, or data 26 selected for high priority transmission through switch 12'. Data 26 thus selected may be bit-wise ANDed with a data mask to obtain revised data. The revised data may be compared with the predetermined data to determine if a participation match exists. Alternatively, the adjusting step also permits selecting frames 18 for which there is no participation match.

As shown best in Figure 4, which shows at least one of a number of ways the present invention may be practiced, each received frame 12 may be presumed to have a quality of service 24 ("QOS") level assigned to it by a receiving port. Each fibre channel switch 12' may provide for a number of separate queues 28, one for each of the receiving ports 30 (exemplified schematically in Figure 4) in a fibre channel switch 12'. Each queue 28 may contain one or more queue entries ("Qentries") 32. Qentries 32 also identify the buffer memory location of each frame 18. Each of the queues 28, as shown in Figure 4, may have a WEIGHT_TIME ("WT") register setting assigned. In addition, the WT register settings may be programmed to provide a greater portion of bandwidth to traffic received from receiving ports 30 with a greater value in the corresponding WT register setting. As used in this document, and as known to those skilled in the art, the term "bandwidth" refers to rate at which a fibre channel system, as exemplified in Figure 2, may transmit data 26, which in turn is based on the range of frequencies that an electronic system can transmit. Each Qentry 32 may be given a bit score, for example an 8 bit score. As provided by the present invention, at least one formulation of the initial score to be assigned is based on the formula *Score = QOS + [WT] [TOQ]*, where the term "TOQ" means time on queue that a Qentry 32 has spent on a queue 28.

As also shown in Figure 4, at least one algorithm that may be used in connection with the practice of the present invention is:

- A Qentry is received from Port_N;
- An initial score for the Qentry is computed from the QOS level assigned by receiving port N, such that Score = QOS;
- The Qentry is inserted into a Queue N ahead of all other Qentries in Queue N with lesser scores, but behind all other Qentries with greater or equal scores;
- Following passage of a selected time period, the scores for every Qentry in Queue N is adjusted by the WT of N; and
- Prior to selection of a frame for transmission, the scores of the Qentries at the bottom of the Queues are compared with the entry with the greatest score for selected transmission through the switch, as best shown in Figure 5.

As shown in Figures 4, at least one example of the determination of the QOS component 24 of the score is shown. As shown, eight (8) templates 34 are programmed to scan data 26 within an incoming frame 18. Each template 34 examines data 26, which may include a word, at a specified location called an "Offset" 36 within incoming frame 18. The data 26 or word is bit-wise

ANDed 38 with a mask word ("Mask") 40, and compared with a content word ("Content") 42. As set out in U.S. Patent 6,233,236, incorporated herein by reference, in a particular example, content 42 includes information extracted from the header fields of each packet transmitted. For example, the header fields of each fibre channel frame include destination ID (i.e., a field identifying the port that is the intended recipient of a frame), a source ID (i.e., a field identifying the port to which the receiver belongs), a frame type (i.e., a field identifying the FC-4 frame type). It should be noted that the destination and source ID information referred to herein refers to intra-switch information and is different from the S_ID and D_ID information in an FC-4 frame that refer to actual fabric device addresses. Other types of information or "metadata" (i.e., data that describes the frame) may be included in the header or other designated fields of a frame depending to the frame format requirements of a particular application. If masked word 40 matches content word 42, a template match has occurred for that template 34. Alternatively, if there is no match, the "Negate" bit 44 is set true, which may be useful if a user seeks to search or scan or for no match, which would be useful, for example, when seeking to define a destination value on frame 18.

Also, a number of participation groups, for example eight (8) "Participation Groups" 46 may be programmed to assign a Quality of Service value 24 to each incoming frame 18. Each Participation Group 46 looks for a participation match for each template 34 corresponding to the bits set to >1= in the Participation field 48 (e. g. if Participation = 0001011 then a participation match occurs if Templates 3, 1, and 0 all have template matches). When a participation match occurs, the Quality of Service level 24 for incoming frame 18 is either assigned the QOS value 24 associated with that participation group 48 (when UseFrameCtl = 0, or to the frame control word from one of the Frame Control word finders when UseFrameCtl = 1 or 2). In addition, while not an essential step in assigning an initial value based on QOS 24, at least two (2) frame control word finders 50 may be programmed to extract a 3-bit value from an incoming frame. Each frame control word finder may extract the

3 least significant bits from a byte in incoming frame 18, and Offset 36 and Byte fields may specify the location of the byte.

The present invention also includes one or more steps for adjusting the initial score. A variety of alternative score components may be applied to the initial score to determine one or more adjusted scores. For example, at least one embodiment of the present invention includes an alternative score component for identifying receiving ports (not shown) for frames 18, and applying a bandwidth allocation adjustment derived from the location pf frame 18 among receiving ports 30. In addition, the present invention provides for calculating the cumulative time that data 26, including frame 18, remain on queue 28 in the transmit switches (not shown) of switch 12' before being transmitted to a receiver device in the fibre channel fabric as suggested by Figure 2. The invention provides for measuring each millisecond (ms) of time data 26 is on queue 28, but different time intervals may be selected. Generally, the initial score is increased by the bandwidth allocation adjustment and for the time on queue.

The present invention also provides for comparing adjusted scores with each other. As best shown in Figure 5, frames 18 having the highest adjusted scores are identified, rearranged in a reordered queue based on the adjusted scores, and frames 18 having the highest adjusted scores may then be transmitted through switches 12.

While the method for scoring queued frames for selective transmission through a switch as shown in drawing figures 1 through 5 is one embodiment of the present invention, it is indeed but one embodiment of the invention, is not intended to be exclusive, and is not a limitation of the present invention. While the particular method for scoring queued frames for selective transmission through a switch as shown and disclosed in detail herein is fully capable of obtaining the objects and providing the advantages stated, this disclosure is merely illustrative of the presently preferred embodiments of the invention, and no limitations are intended in connection with the details of construction, design or composition other than as provided and described in the appended claims.